XML Approach for the Solution of Chain of Custody of Digital Evidence

Devi Ratnasari
Department of Informatics
Universitas Teknologi Yogyakarta Yogyakarta Indonesia

Yudi Prayudi
Department of Informatics
Universitas Islam Indonesia
Yogyakarta Indonesia

Bambang Sugiantoro
Department of Informatics
UIN Sunan Kalijaga
Yogyarta Indonesia

ABSTRACT
Traditionally, the concept of recording information about physical evidence (electronic and non-electronic) in various law enforcement agencies is done using paper-based documentation. A form is generally applied to document the chronological information of the evidence during the investigation process. In the future, the traditional concept can be problematic when it applied to digital evidence. It is because the physical evidence and the digital evidence have different characteristics.

This paper present the Chain of Custody application for digital evidence documentation using XML approach. XML is selected as the chain of custody information documentation storage media. In the application, there are two types of the chain of custody information, such as information entered by the user and information extracted from the attribute of digital evidence file. This application can be applied for various file types from various electronic sources. In addition, this paper also conducted a preliminary experiment using this application to know that Chain of custody information can be properly recorded and this approach does not alter the digital evidence. The output of this application is the chain of custody form in the .pdf document format.

General Terms
Digital Forensics, Chain Of Custody.

Keywords
Cyber Crime, Digital Evidence, Digital Chain Of Custody, XML Schema

1. INTRODUCTION
During the investigation process, one of the important steps is to ensure that the evidence is handled properly by the responsible party. Evidence for the investigation process of cybercrime is divided into two types: physical evidence (electronic) and digital evidence. Physical evidence (electronic) is an electronic device that can be used to aim the cybercrime activity or other devices that can record the trace of cybercrime activity. The examples of physical evidence are harddisk, CD, pen drive, CCTV, computer, RAM, mobile phone and etc. While digital evidence is the digital content of the acquisition and extraction results from the physical evidence (electronic). Digital evidence can also a digital content bit by bit data extraction (full copy) of storage media such as hard drives, flash drives, floppy disks and optical media. This digital evidence is known as Disk Image file [1] [2].

Because the chain of custody documentation still does not have a global standard, every law enforcement in a various country can have different form of the chain of custody documentation. So far there has been no regulation or standard rules that become the main reference for the organization in conducting activities and determine the needs of the chain of custody information, especially for digital evidence. However to be accepted in court, the custody form should include at least "5W and 1H" information about What, Who, Where, When, Why and How according to the evidence. Those pieces of information such as the person who involved in handles of the evidence, the time of each process undertaken, methods used on how the process was carried out, displacement of the evidence and storage location, the reason why the party handles the evidence and what evidence has been collected [3].

In various agencies, the mechanism of implementation of the chain of custody documentation for physical evidence has been done using paper-based such as forms, control books or register books. This mechanism would be difficult if applied to document the chain of custody of digital evidence. It is because the physical evidence and digital evidence have different characteristics. But for the concept and information needed in the chain of custody documentation between those two should be the same [4].

Given the importance of a chain of custody solution for digital evidence in cybercrime investigation process, a system-based/application-based for the chain of custody solution is extremely important. This paper presents an application architecture and illustrates on how the implementation using an XML (Extensible Markup Language) schema approach. The XML concept preserves the integrity of the hash value of evidence files so that the evidence can be well documented and acceptable in court.

2. CURRENT ISSUE
Chain of custody is a procedure for evidence chronological record documentation. The record starts from the time when evidence was found, the process of duplication and examination, the storage of evidence (physical or digital) to the presentation and final decision or disposal of the evidence. Chain of custody can also be used to ensure the integrity and originality of the evidence [5].

The documentation of the chain of custody has been widely applied to physical evidence. While for digital evidence there are a number of solutions provided by other researchers such as the business model of digital evidence handling [4], the creation of a digital proofread chain of custody framework with the digital cabinet concept [6], the framework of the chain of custody process of digital evidence investigation [7], the description of the information needs of chain of custody

management using the ontology approach [8] [9], the concept of digital evidence bag that mimics the function of a plastic bag of evidence by using an XML document type [10] and others.

Based on some of the above research, the existing solutions are still unable to fulfill the requirement for the chain of custody documentation for digital evidence. Therefore, a solution is needed on how technically the chain of custody documentation can be performed for digital evidence.

The Chain of Custody application for digital evidence using the XML schema approach comes as one of the proposed solutions to enrich the existing solution of the digital chain of custody.

3. DESIGN ARCHITECTURE SYSTEM

The Digital Chain of Custody application is an implementation prototype using the XML schema approach. The XML schema is used as a means to store the chain of custody information. Some advantages of using an XML schema are it's strong, durable, easy to manipulate, data independence, interoperability and free for anyone. XML-based is also easy for identify, store and transfer of information [11].

The Digital Chain Of Custody application can be used to document digital evidence files from various sources including forensic hard disk images, network packet captures, any of document files, multimedia files and more. The main functions accommodated by the application are extract the attribute information of digital evidence file, create documentation files, modify documentation information and save the documentation information on the application. Those functions can be shown in Digital Chain of Custody application architecture in Figure 1.

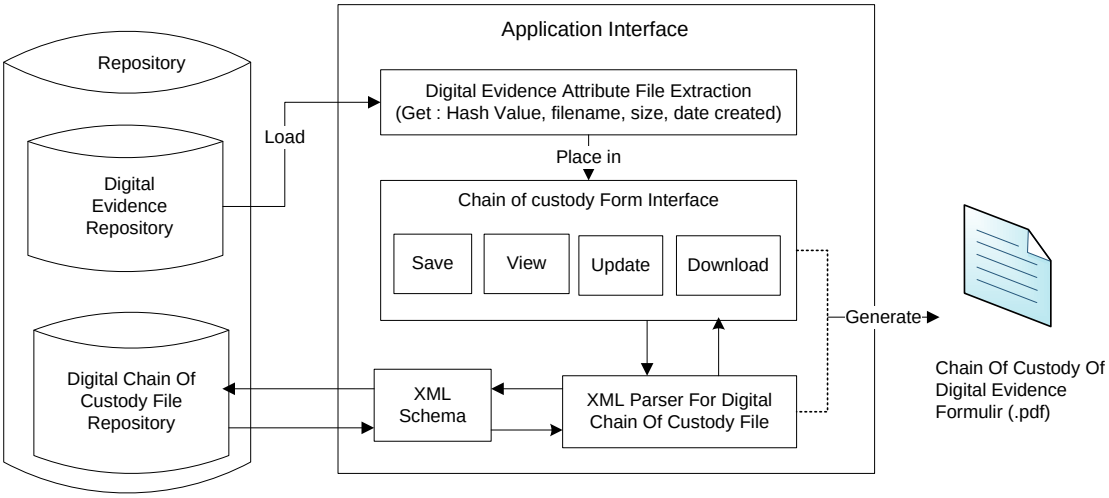

Fig 1: Digital chain of custody application architecture

This proposed application architecture is basically designed to support various types of digital evidence file such as extracted document files, multimedia files, disk image files, network packet capture files and more. The type of the files also includes from various electronic evidence sources such as a computer, mobile phone, USB drive, laptop, and others. The process or method of acquisition to produce the digital evidence file are also from an online and offline acquisition. This architecture was inspired based on a digital forensic handling business practice model on digital evidence using a digital cabinet approach as a repository of digital evidence file [6].

Broadly speaking this application has three main concept segments that establish the needs for the chain of custody of digital evidence can be fulfilled. The segments include the concept of a repository, the concept of recording the activity flow and the field of information used in the chain of custody document.

3.1 Concept Of Repository

This application stores digital evidence files and chain of custody files separately in the main repository. The main repository is divided into two parts, there are digital evidence repository and digital chain of custody file repository. Digital evidence repository is a directory containing the digital evidence files. The digital chain of custody file repository is a directory containing the result files from the chain of custody information entry of digital evidence with .xml documents format. For more details can be shown in the Figure 2:

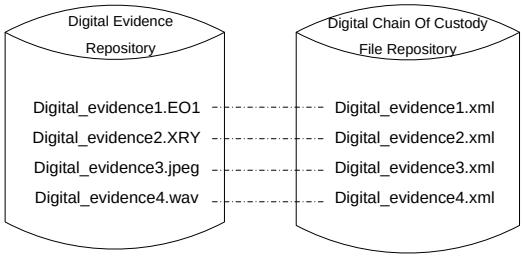

Fig 2: Application repositoty concept

Practically, one digital evidence file will have only one chain of custody file with XML format. Reciprocally one XML chain of custody file will only be documenting one digital evidence file. The name of the chain of custody file is also customized with the name of the digital evidence file. For example, if the name of the digital evidence file is "recording.amr" then the chain of custody file name will be "recording.xml".

3.2 Flow Of Recording

The chain of custody information recording activity is performed using a form within the interface of application. The recording of the chain of custody information starts from

the electronic evidence, the acquisition process by First Responder and the process of storing the acquisition results into storage media (repository). The officer is a person who responsible for the chain of custody information and check the evidence in and out from the storage facility. This application consists of two users, there are the user as a first responder and the user as an officer. Before doing any activity inside this application, the user will be prompted to log in. And these two users will certainly have different permissions in carrying out the chain of custody information documentation activities. User as an officer has full control of information and application. Therefore, every activity performed by the first responder must obtain approval from the officer. The chain of custody information entered by the first responder will also be checked first to be validated before being stored in the digital chain of custody file repository.

This application interface contains 3 main menus there a file, manage and view. The recording functions accommodated within the application are load the digital evidence file into the application, extract the file attribute information, entry and save the new chain of custody information, modify the existing chain of custody information and generate a report from the chain of custody information presented in a form with a .pdf document.

To perform the act of creating the Chain of Custody documentation file in Digital Chain of Custody application can be done in accordance with the flowchart shown in Figure 3.

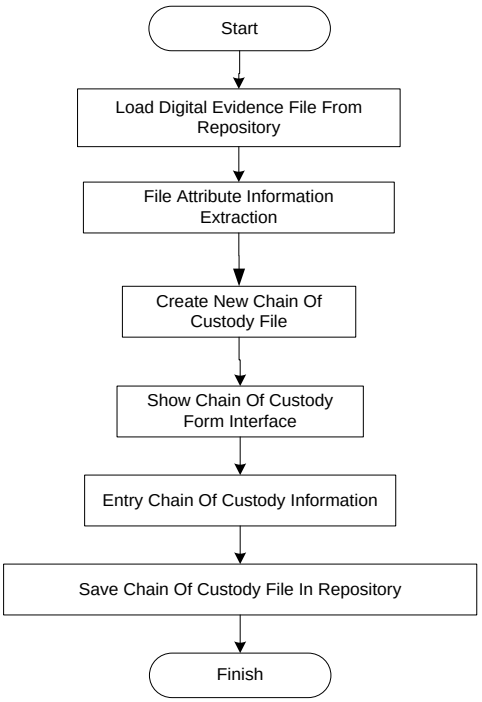

Fig 3: Flowchart of create chain of custody file

Whereas to perform modification (ie. add, change and delete) activities, information documentation of Chain of Custody file can be done in accordance with the flowchart shown in Figure 4.

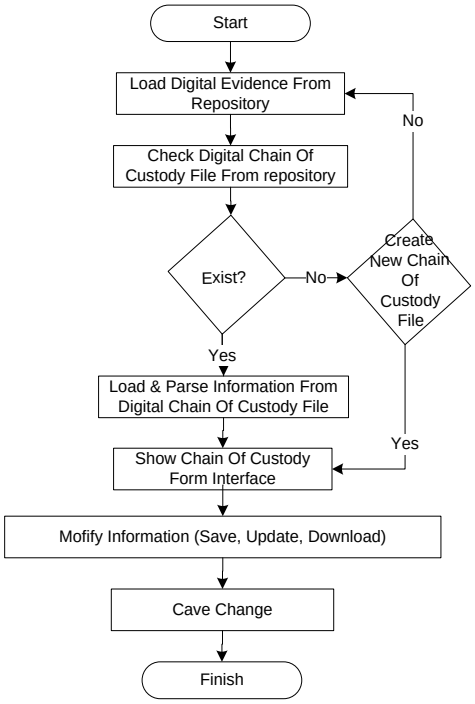

Fig 4: FLowchart of modify chain of custody file

When the user modifies the chain of custody information by loading a digital evidence file from a digital evidence repository, the application checks the chain of custody file in the digital chain of custody repository. If the chain of custody file was not found, the application will display a blank form and ask the user to create a new chain of custody file. However, if the chain of custody file was found, the application will display the chain of custody information in the form according to the appropriate field.

Every chain of custody information from creating and modifying activity will be stored in the repository with the XML schema. The schema stores each information field inside the XML tag. In order to store and access information from XML tags, the application requires a library to parse information. In this research the application using JDOM parser library. Conceptually, the information entered by the user in the field will be parsed by JDOM and stored in the XML tag. Likewise, JDOM will parse any tag information from an XML file if the user needs to read the information and display it into the application.

3.3 Chain Of Custody Information Fields

The information of the chain of custody of digital evidence will only be changed or updated in case of interaction with the evidence from and to the storage facility (Digital Evidence Repository). The information fields for the chain of custody are obtained by identifying and extracting from some chain of custody form model downloaded from the internet. The form that used in identification and extraction is a form of evidence chain of custody for a computer crime case. Despite having the same goal, but this form has characteristics and differences. The differences can be seen from the type of evidence and the number of items of the evidence that can be accommodated in a single chain of custody form, the information stored and the number of information fields provided in the form.

Among those forms that used to extract the information fields are the chain of custody form from University of

Pennsylvania, West Audit, NIST (National Institute of Standards and Technology), Digital Forensic Lab and PVL Forensics.

The normalization has been done by removing some fields that are less appropriate and add some fields that are needed for the chain of custody information on the application. The purpose is that the chain of custody information needs for digital evidence can be achieved. From the normalization process, there are 42 information fields for the chain of custody of digital evidence. Some information fields are selected from existing forms and some are additional information fields.

4. IMPLEMENTATION & RESULT

The Digital Chain Of Custody application was built using Java language and cross-platform compliant. Implementation of the application architecture and the 42 chain of custody information fields were created like physical form but actually it is in digital form. If inside the physical form usually the information grouped in a particular section, then in the application the information were organized and grouped in several tabs of information. There are collection information, electronic evidence information, digital evidence information and interaction information. Some experiment using this application have been conducted for the chain of custody documentation of the digital evidence file. Here are some screenshots that show the chain of custody application.

Figure 5 is a form page which displays the chain of custody information entry. This page will appear after the user loads a digital evidence file into the application. The application form consists of five tabs. Each tab has a different information field. This concept is like information division on a physical form. These pictures show information on the digital evidence information tab. It has been mentioned before that one of the applications function is able to extract the attribute information of digital evidence file. The extracted information was then included in the relevance information field in the application form. Among the attribute information file in the digital evidence tab such as filename and file type, size of the file, date created and time created. The application will also compute hash values from digital evidence files (MD5, SHA1, and SHA-256). Since the application computes three hash values, the load speed, the extraction of the attribute information and the digital hash value then depends on how big the file is.

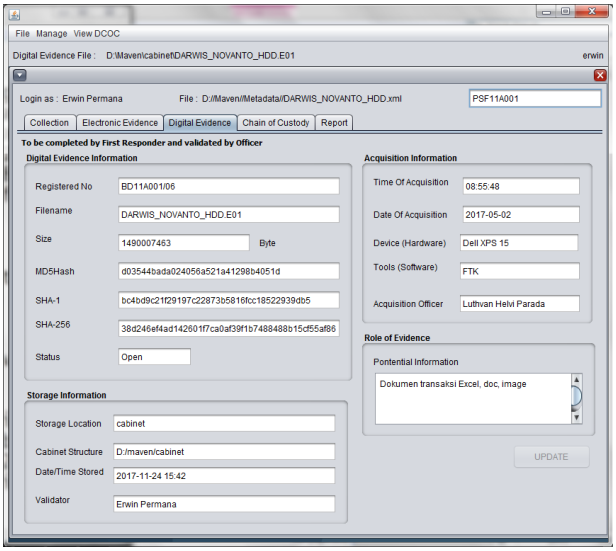

Fig 5: Interface of digital evidence information tab

Figure 6 shows the form page of the chain of custody tab. This section is used to input interaction information against digital evidence. Chain of custody tab can only be accessed by a user responsible for checking in and out of evidence from the storage facility. In this case, the user is an officer.

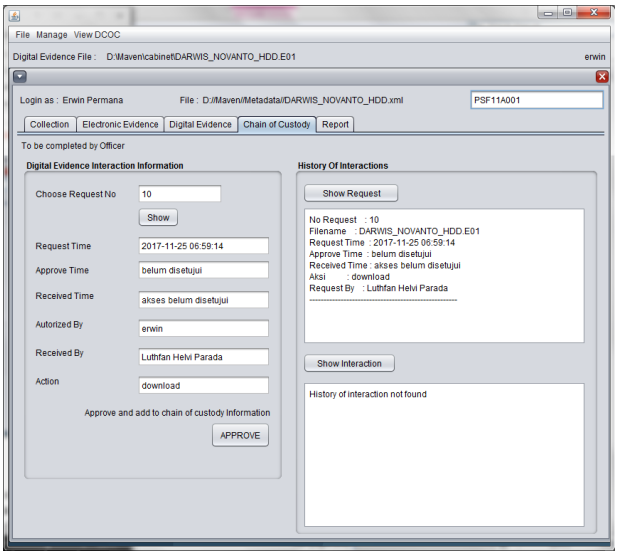

Fig 6: Interface of chain of interactions information tab

In addition to the act of entering the chain of custody information, this application also provides a function to generate form reports in .pdf format documents. The generate report function of the digital chain of custody form is on the Reports Tab.

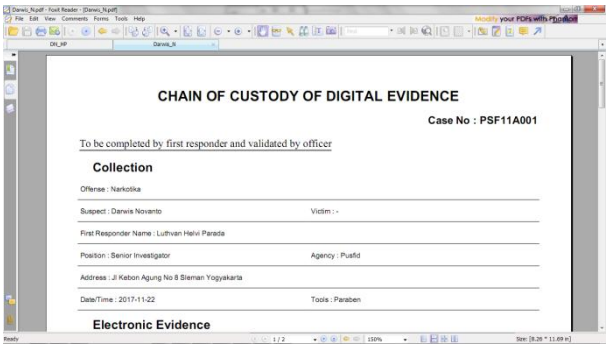

Fig 7: Digital chain of custody report

Based on the design of the chain of custody form in Fig. 8, the result file of the chain of custody information of the 42 information fields on the form is stored inside the XML tag. The XML tags schema file documentation of the chain of custody is as follows:

XML Tag schema of chain of custody file

```
<?xml version="1.0" encoding="UTF-8"?>
<chain_of_custody>
  <case_information>
    <case_no meta="dinamik"></case_no>
    <offense meta="dinamik"></offense>
    <suspect meta="dinamik"></suspect>
    <victim meta="dinamik"></victim>
  </case_information>
  <first_responder>
    <first_responder_name
meta="dinamik"></first_responder_name>
    <agency meta="dinamik"></agency>
    <position meta="dinamik"></position>
```

```
  </first_responder>
....
....
</chain_of_custody>
```

5. DISCUSSION

The application implementation analysis is done by an experiment using a case scenario involving multiple digital evidence files. This experiment will let us know that a chain of custody approach to digital evidence using XML schemas will not reduce or change the integrity value of digital evidence files. The value of the integrity of digital evidence can be shown by the changed of MD5 hash value between the digital evidence file before and after the chain of custody documentation.

The data of the implementation in this study used a computer or cybercrime case scenario. There are four pieces of evidence found during the collection process in a crime scene. The pieces of evidence are two files of direct evidence reported to the officer and electronic evidence such as a computer, USB flash drive, smartphone, and desktop. There are two types of acquisition methods used in the investigation: online acquisition and offline acquisition. The tools used during the acquisition process are also made from various forensic software such as Encase, FTK, XRY and Wireshark network packet capture software. The MD5 hash value of the digital file has also been calculated before and after carrying out the process of entering the chain of custody information using the application.

From the case of scenario implementation process using the chain of custody application for 6 digital evidence files produce 6 chain of custody files in .xml format. The result of the chain of custody filename is in accordance with the name of the digital evidence file. The file name is given automatically by the application to avoid the user for making a mistake when entering the file name. Each of the chains of custody files contains chronological information from a single digital evidence file. Each of the chains of custody file can also represent a manual form of the chain of custody information but stored in an XML tag. That information stored in XML tag then will be created in .pdf report format and will generate a document as same as a manual form design. One of the examples is shown in Figure 7, this document is a form document generated from the chain of custody information of the digital evidence file namely "DARWIS_NOVANTO_HDD.E01". Each field of information contained in the document form is available in the XML schema of the chain of custody file namely "DARWIS_NOVANTO_HDD. XML".

Below is a snippet of the tag information from an XML schema of a chain of custody file presented from the "DARWIS_NOVANTO_HDD.xml" file :

Tag XML Schena, Of "DARWIS_NOVANTO_HDD.xml file

```
<?xml version="1.0" encoding="UTF-8"?>
<chain_of_custody>
  <case_information>
    <case_no meta="dinamik">PSF11A001</case_no>
    <offense meta="dinamik">Narkotika</offense>
    <suspect meta="dinamik">Darwis
Novanto</suspect>
    <victim meta="dinamik">-</victim>
  </case_information>
  <first_responder>
.
.
<digital_evidence>
    <digital_evidence_no
meta="dinamik">BD11A001/06</digital_evidence_no
>
    <filename
meta="statik">DARWIS_NOVANTO_HDD.E01</filename>
    <size meta="statik">1490007463</size>
    <md5
meta="statik">d03544bada024056a521a41298b4051d<
/md5>
    <sha1
meta="statik">bc4bd9c21f29197c22873b5816fcc1852
2939db5</sha1>
    <sha256
meta="statik">38d246ef4ad142601f7ca0af39f1b7488
488b15cf55af8613f7a575f64b663d</sha256>
.
.
</chain_of_custody>
```

The results of the acquisition process undertaken, the chain of custody information entry using the application and hash value before and after the entry of chain of custody information can be seen in Table 1.

Table 1. Result of experiment : chain of custody of digital evidence using cybercrime case scenario

No	Electronic Evidence	Acquisition and Disk Imaging	Tools	Digital Evidence	Integrity MD5 Before Documentation	Digital Chain of Custody	Integrity MD5 After Documentation
1	Harddisk	Offline	Encase	DARWIS_NOVANTO_HDD.E01	d03544bada024056a521a41298b4051d	DARWIS_NOVANTO_HDD.E01	d03544bada024056a521a41298b4051d
2	Flash drive	Offline	FTK	DN_HP2_FDD.dd	5f800f3cf7660885fccb1cf012a6b29f	DN_HP2_FDD.xml	5f800f3cf7660885fccb1cf012a6b29f
3	Smartphone	Offline	XRY	DarwisN_samsung.xry	2e6a6dac2195eab1e1ed2e8fc957dfbe	DarwisN_samsung.xml	2e6a6dac2195eab1e1ed2e8fc957dfbe
4	Desktop	Online	Wireshark	DarwisN_cap1.pcap	40e7e81ae206f291b531c9252132d7ab	DarwisN_cap1.xml	40e7e81ae206f291b531c9252132d7ab
5		-	-	Selfie.jpg	195d5e069d225ebb7adf565a0cefd18a	Selfie.xml	195d5e069d225ebb7adf565a0cefd18a
6		-	-	Afwan_Video1.mp4	df6133f4b68a12ebf3a88db51a679abe	Afwan_Video1.xml	df6133f4b68a12ebf3a88db51a679abe

6. CONCLUSION & FUTURE WORK

A design of application architecture has been proposed as one of the solutions for digital evidence documentation. The XML schema approach is used as a tool for storing the chain of custody information. The digital evidence files and chain of

custody files are stored in separate repositories. The application automatically extracts digital file attribute information that is used as a chain of custody information and links between a digital evidence file and a chain of custody file. Experiments have been performed on several types of digital evidence files using a cybercrime case scenario. Based on the implementation and analysis, the chain of custody application using XML schema approach can be used to document the chain of custody information of digital evidence file quite well. Furthermore, this application does not change the MD5 value or the integrity value of digital evidence file based on the hash value of the digital evidence file before and after the entry of chain of custody information.

In the future work, this chain of custody concept using the XML schema approach can be developed and applied to document the information for the chain of custody of digital evidence. One of the limitations of the application in the documentation of digital evidence information is that this application still retrieves information from the digital file attribute value. This application will be better and more accurate if the digital evidence information can be extracted from the metadata information of the digital evidence file. This paper also did not pay attention to the context of the information fields required in the chain of custody information documentation for digital evidence. The further studies related to the application of access control and the need for information fields for the chain of custody form are also indispensable.

7. REFERENCES

[1] B. Carrier, FileSystem Forensic Analysis. Adisson Wesley Professional, 2005.

[2] Y. Prayudi, "Problems and Solutions for Digital Chain Of Custody in Cybercrime Investigation Process (In Indonesian Language)" Senasti, no. ISSN : 235-536X, 2014.

[3] J. Cosic, "Formal Acceptability of Digital Evidence," Springer Int. Publ., 2017.

[4] A. Luthfi and Y. Prayudi, "Digital Forensics Business Model for Supporting of Handling Digital Evidence and Cybercrime Investigation (In Indonesian Language)" Konf. Nas. Inform. STIE ITB Bandung, 2015.

[5] Y. Prayudi and A. SN, "Digital Chain of Custody : State of the Art," Int. J. Comput. Appl., vol. 114, no. 5, p. 8887, 2015.

[6] Y. Prayudi, A. Ashari, and T. K. Priyambodo, "Digital Evidence Cabinets : A Proposed Framework for Handling Digital Chain of Custody," Int. J. Comput. Appl., vol. 107, no. 9, pp. 30–36, 2014.

[7] J. Cosic and M. Baca, "A Framework to (Im) Prove „ Chain of Custody " in Digital Investigation Process," Proc. 21st Cent. Eur. Conf. Inf. Intell. Syst., pp. 435–438, 2010.

[8] Y. Prayudi, A. Luthfi, A. Munasir, R. Pratama, and K. Kunci, "An Ontological Approach for Representing Body of Knowledge of Digital Chain of Custody (In Indonesian Language)" Cybermatika, vol. 2, pp. 36–43, 2014.

[9] J. Cosic, Z. Cosic, and M. Baca, "An Ontological Approach to Study and Manage Digital Chain of Custody of Digital Evidence," JIOS J. Inf. Organ. Sci., vol. 35, no. 1, pp. 1–13, 2011.

[10] C. Hosmer, "DIGITAL EVIDENCE BAG," Commun. ACM, vol. 49, no. 2, 2006.

[11] S. Airi and F. Tompa, "'Why Use XML ?'. In Communicating with XML," Springer US, pp. 69–91, 2011.